

ALGORITMA C4.5 UNTUK MENENTUKAN TINGKAT KELAYAKAN MOTOR BEKAS YANG AKAN DIJUAL

Dede Wira Trise Putra

Dosen Jurusan Teknik Informatika

Fakultas Teknologi Industri

Institut Teknologi Padang

Email : dedewtp339@yahoo.com

Abstrak

Pohon keputusan merupakan representasi sederhana dari teknik klasifikasi yang merupakan proses pembelajaran suatu fungsi tujuan yang memetakan tiap himpunan atribut ke satu dari kelas yang didefinisikan sebelumnya. Implementasi algoritma C4.5 dalam menentukan tingkat kelayakan motor bekas yang akan dijual untuk membantu proses pengklasifikasian kondisi motor bekas. Pohon keputusan dapat menemukan hubungan tersembunyi antara sejumlah variabel input dengan sebuah variabel target dari data penjualan motor bekas. Pada algoritma C4.5 dilakukan penghitungan entropy dan gain information untuk memperoleh node akar dan node lainnya. Dengan kemampuannya untuk mem-break down proses pengambilan keputusan yang kompleks menjadi lebih mudah. Pohon keputusan yang dihasilkan dari kasus yang diangkat menunjukkan bahwa ada beberapa atribut yang mempengaruhi dalam penentuan kelayakan motor bekas yakni aki, mesin, bodi, cat dan aksesoris.

Kata Kunci : data mining, C4.5, pohon keputusan, kelayakan

Abstract

Decision tree is a simple representation of a classification technique that is a learning process an objective function that maps each set of attributes to one of the classes defined previously. C4.5 algorithm implementation in determining the feasibility of a used motorcycle that will be sold to help the classification process conditions used motorcycles. Decision trees can discover the hidden relationship between the number of input variables with a target variable of the data used bike sales. At C4.5 algorithms were calculated entropy and gain information to obtain root node and other nodes. With its ability to break down complex decision-making process becomes easier. The resulting decision tree of the cases raised shows that there are some attributes that affect the determination of the feasibility of a used motorcycle battery, engine, body, paint and accessories.

Keywords : data mining, C4.5, decision trees, feasibility

1. PENDAHULUAN

1.1. Latar Belakang

Komputer sebagai alat bantu pekerjaan manusia saat ini mengalami perkembangan yang sangat pesat. Perkembangan tersebut tidak hanya terjadi pada satu bidang, namun pada banyak bidang. Salah satu perkembangan yang saat ini sangat dirasakan manfaatnya adalah penemuan informasi baru dengan mencari pola baru dari suatu data yang sangat besar yang dikenal dengan istilah Data Mining. Data Mining merupakan proses dalam menganalisis data dari sudut pandang yang berbeda dan membentuknya menjadi informasi yang berguna sehingga

pengguna dapat memahami substansi hubungan antara data.

Data Mining memiliki beberapa metode yang salah satunya ialah metode klasifikasi merupakan teknik Data Mining yang memetakan data ke kelompok yang telah ditetapkan. Metode klasifikasi dianggap penting karena metode ini dapat memudahkan pengelompokan terhadap data yang akan diolah.

Penggunaan metode klasifikasi dapat memanfaatkan beberapa algoritma di antaranya Algoritma C4.5. Dalam algoritma C4.5 terjadi proses pengklasifikasian data dalam bentuk tingkatan yang berbeda mulai dari akar ke daun

secara hirarki. Proses ini berlanjut hingga mencapai node yang tidak dapat dibagi lagi.

Proses yang terjadi dalam pembentukan pohon keputusan adalah mengubah bentuk data menjadi model pohon, mengubah model pohon menjadi *rule*, dan menyederhanakan *rule* yang ada. Dalam algoritma C4.5 nantinya akan diambil informasi dari data-data yang pasti dalam membantu pemakai dalam mengambil keputusan dari masalah ini.

1.2. Perumusan Masalah

Berdasarkan latar belakang yang diuraikan di atas, maka permasalahan yang akan dibahas antara lain :

1. Bagaimana Data Mining Algoritma C4.5 dapat diterapkan untuk menentukan kelayakan motor bekas yang akan dijual tersebut?
2. Bagaimana Data Mining Algoritma C4.5 membantu Sun Motor dalam pengambilan keputusan dalam pengadaan motor-motor yang akan dijual?

1.3. Batasan Masalah

Agar penulisan ini dapat terarah dan sesuai dengan tujuan yang akan dicapai, maka penulis membatasi penelitian sebagai berikut :

1. Penelitian yang penulis lakukan hanya untuk menganalisa penggunaan algoritma C4.5 dalam penentuan kelayakan motor yang akan dijual.
2. Metode yang digunakan klasifikasi dengan menggunakan algoritma C4.5 untuk membuat pohon keputusan.
3. Aplikasi yang penulis gunakan untuk menguji hasil penelitian adalah *Rapid Miner*.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk :

1. Mengetahui dan mengimplementasikan proses penentuan kelayakan motor bekas dengan menggunakan algoritma C4.5.
2. Menganalisa kelayakan motor bekas menggunakan algoritma C4.5.
3. Mengevaluasi *rule* dengan membuat pohon keputusan.
4. Menjelaskan proses pembentukan pohon keputusan untuk menentukan kelayakan

motor yang akan dijual dengan metode algoritma C4.5.

5. Merancang pohon keputusan untuk menentukan kelayakan motor bekas yang akan dijual.
6. Membangun aplikasi untuk menciptakan pohon keputusan.
7. Menguji data yang diperoleh dengan dengan aplikasi yang dibangun untuk membentuk pohon keputusan.

1.5. Manfaat Penelitian

Penelitian ini memiliki manfaat sebagai berikut :

1. Memberikan kemudahan menentukan harga penjualan motor bekas.
2. Menjadikan pedoman untuk menghasilkan motor bekas yang akan dijual dalam kondisi layak.
3. Memberi masukan dalam menentukan kriteria kelayakan motor.

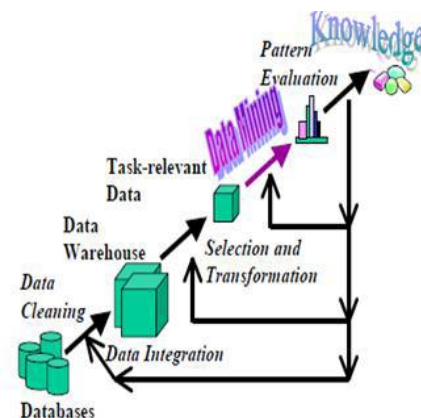
2. METODOLOGI

2.1. Knowledge Discovery in Databases (KDD)

Knowledge Discovery in Databases (KDD) merupakan suatu proses yang validasi, manfaat, dan pola yang bisa dipahami dalam sebuah data yang tersimpan dalam sebuah *database* yang besar. Proses dalam KDD yakni seleksi, pra-pengolahan, transformasi, data mining, dan interpretasi.

2.2. Data Mining

Data Mining merupakan suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Proses Data Mining terlihat seperti Gambar 1.



Gambar 1. Proses dalam Data Mining

2.3. Algoritma C4.5

Pada tahap belajar dari data, algoritma C4.5 mengkonstruksi pohon keputusan dari data pelatihan, yang berupa kasus-kasus atau *record-record (tupel)* dalam *database*. Tiga prinsip kerja algoritma C4.5 pada tahap belajar dari data adalah :

1. Pembuatan pohon keputusan.
2. Pemangkas pohon keputusan dan evaluasi (opsional).
3. Pembuatan aturan-aturan dari pohon keputusan (opsional).

Rumus pada algoritma C4.5 adalah :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * Entropy(S_i) \quad \dots(1)$$

Di mana :

S : himpunan kasus
A : atribut

n : jumlah partisi atribut A

$|S_i|$: jumlah kasus pada partisi ke- i

$|S|$: jumlah kasus dalam S

$$Entropy(S) = \sum_{i=1}^n \left| \frac{S_i}{S} \right| - p_i * \log_2 p_i \quad \dots(2)$$

Di mana :

S : himpunan kasus

A : fitur

n : jumlah partisi S

p_i : proporsi dari S_i terhadap S

3. HASIL DAN PEMBAHASAN

3.1 Analisa Data Mining dengan Algoritma C4.5

Proses pengambilan keputusan dalam klasifikasi kelayakan motor bekas yang akan dijual dikelompokkan menjadi beberapa kriteria yaitu :

Tabel 1. Format Data Akhir

No. Polisi	Jenis Motor	Mesin	Rangka	Body	Cat	Aki	Aksesoris	Keputusan
5017 WU	BLADE	halus	mulus	lecet	baret	soak	semi modif	perbaikan
4102 WU	ABSOLUTE REVO SPOKE	halus	mulus	mulus	kusam	normal	ori	langsung jual
4819 WJ	REVO CW	Halus	keropos	lecet	ganti cat	soak	semi modif	perbaikan
4743 WH	REVO CW	halus	karat	lecet	kusam	normal	ori	langsung jual
4540 WC	REVO STANDART	halus	mulus	lecet	baret	normal	ori	langsung jual
5148 WQ	BLADE	halus	karat	mulus	kusam	normal	semi modif	perbaikan
6183 HU	ABSOLUTE REVO CW	halus	mulus	lecet	baret	normal	ori	langsung jual
5569 WO	REVO STANDART	halus	karat	mulus	mulus	normal	ori	langsung jual
4005 BP	FIT X	kasar	karat	penyok	ganti cat	normal	ori	perbaikan
4929 WS	ABSOLUTE REVO SPOKE	halus	karat	mulus	mulus	normal	ori	langsung jual
6412 WU	ABSOLUTE REVO SPOKE	halus	mulus	mulus	mulus	normal	ori	langsung jual
6019 BQ	ABSOLUTE REVO CW	halus	mulus	lecet	baret	normal	ori	langsung jual
6383 GW	FIT X	kasar	mulus	mulus	mulus	normal	ori	perbaikan
6920 DT	ABSOLUTE REVO SPOKE	halus	mulus	lecet	baret	normal	ori	langsung jual
6162 WU	ABSOLUTE REVO SPOKE	halus	mulus	mulus	mulus	soak	ori	perbaikan
4865 BP	FIT X	halus	karat	mulus	mulus	soak	semi modif	perbaikan
4957 BR	ABSOLUTE REVO SPOKE	halus	mulus	lecet	baret	normal	ori	langsung jual
6509 BQ	ABSOLUTE REVO SPOKE	halus	mulus	mulus	mulus	normal	semi modif	langsung jual
4540 GX	BLADE	halus	mulus	penyok	ganti cat	normal	semi modif	perbaikan
4384 US	ABSOLUTE REVO SPOKE	halus	mulus	lecet	baret	soak	semi modif	perbaikan
4241 GX	ABSOLUTE REVO SPOKE	kasar	mulus	mulus	mulus	normal	full modif	perbaikan
5105 FS	NEW SUPRA FIT T	halus	karat	penyok	ganti cat	soak	ori	perbaikan
4983 BR	BEAT	kasar	karat	mulus	kusam	normal	semi modif	perbaikan

No. Polisi	Jenis Motor	Mesin	Rangka	Body	Cat	Aki	Aksesoris	Keputusan
6621 FY	FIT X	kasar	mulus	mulus	kusam	soak	semi modif	perbaikan
6792 BN	FIT S	kasar	keropos	lecet	baret	soak	full modif	perbaikan
5281 BR	ABSOLUTE REVO SPOKE	halus	keropos	lecet	baret	soak	ori	perbaikan
4340 WF	REVO CW	kasar	mulus	penyok	ganti cat	soak	semi modif	perbaikan
5175 WF	REVO STANDART	halus	keropos	penyok	ganti cat	soak	ori	perbaikan
4098 WU	BLADE	halus	mulus	mulus	mulus	normal	semi modif	langsung jual
4627 WW	ABSOLUTE REVO CW	halus	mulus	lecet	kusam	normal	ori	langsung jual
5572 WU	BEAT	halus	mulus	lecet	baret	normal	semi modif	langsung jual
6869 AR	ABSOLUTE REVO CW	halus	mulus	mulus	kusam	soak	ori	perbaikan
5798 WS	ABSOLUTE REVO DELUXE	halus	mulus	mulus	kusam	normal	semi modif	perbaikan
6924 WS	ABSOLUTE REVO SPOKE	halus	mulus	lecet	kusam	normal	ori	langsung jual
5387 WS	ABSOLUTE REVO SPOKE	halus	mulus	lecet	kusam	normal	ori	langsung jual

Tahapan algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut :

1. Pilih atribut sebagai akar.
2. Buat cabang untuk tiap-tiap nilai.
3. Bagi kasus dalam cabang.
4. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

3.2. Pohon Keputusan Hasil Perhitungan

Berdasarkan tahapan dalam algoritma C4.5 maka diperoleh hasil seperti yang terlihat pada penjelasan di bawah ini.

Entropy (Aki, Normal and Bodi, Mulus and Mesin, Halus)

$$Entropy(total) = \left(-\frac{2}{3} * \log_2\left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} * \log_2\left(\frac{1}{3}\right)\right) = 0.918296 \quad \dots\dots\dots(3)$$

Entropy(total) adalah menghitung nilai total keputusan perbaikan (2) dan langsung jual (1), sedangkan 3 adalah jumlah keseluruhan kasus.

a. Atribut Rangka

$$Entropy(karat) = \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) = 0$$

$$Entropy(keropos) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(mulus) = \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} * \log_2\left(\frac{1}{2}\right)\right) = 1.000000 \quad \dots\dots\dots(4)$$

Menghitung nilai *Entropy* atribut rangka berdasarkan dari tiap-tiap kelas (karat, keropos dan mulus) pada atribut rangka.

b. Atribut Aksesoris

$$Entropy(full modif) = \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) + \left(-\frac{0}{0} * \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$Entropy(ori) = \left(-\frac{0}{1} * \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} * \log_2\left(\frac{1}{1}\right)\right) = 0$$

$$Entropy(semi modif) = \left(-\frac{2}{2} * \log_2\left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2} * \log_2\left(\frac{0}{2}\right)\right) = 0 \quad \dots\dots\dots(5)$$

Menghitung nilai *Entropy* aksesoris berdasarkan atribut dari tiap-tiap kelas (full modif, ori dan semi modif) pada atribut aksesoris.

Menghitung Nilai *Gain* tiap-tiap atribut menggunakan rumus (1):

Menghitung Nilai *Gain* tiap-tiap atribut *Gain* (Total, Rangka)

$$= Entropy(S) - \sum_{i=1}^n \frac{|Rangka_i|}{|Total|} * Entropy(Rangka_i) \quad \dots\dots\dots(6)$$

$$= 0.918296 - \left(\left(\frac{1}{3} * 0\right) + \left(\frac{0}{3} * 0\right) + \left(\frac{2}{3} * 1\right)\right)$$

$$= 0.251629$$

$Gain(\text{Total, Aksesoris})$

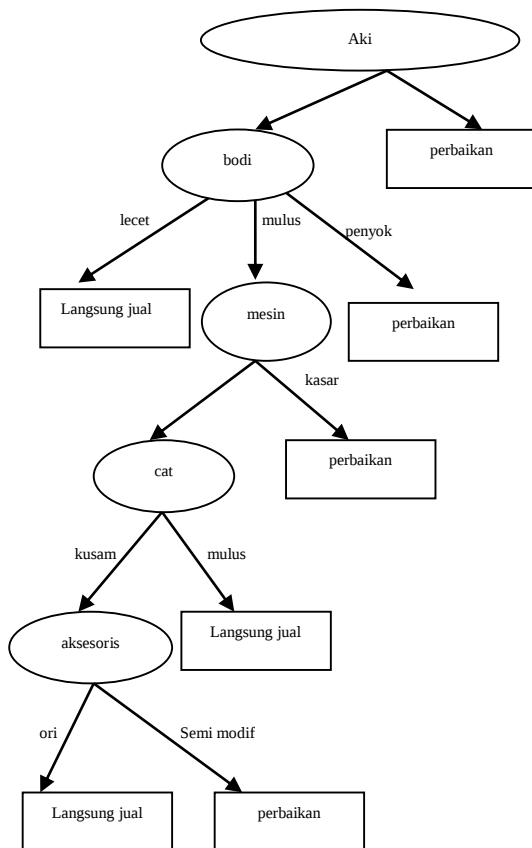
$$= Entropy(S) - \sum_{i=1}^n \frac{|Aksesoris_i|}{|Total|} * Entropy(Aksesoris_i) \dots\dots(7)$$

$$= 0.918296 - \left(\left(\frac{0}{3} * 0 \right) + \left(\frac{1}{3} * 0 \right) + \left(\frac{2}{3} * 0 \right) \right)$$

$$= 0.918296$$

Tabel 2. Perhitungan Node 1.4

Node		Jumlah Kasus (S)	Perbaikan (S1)	Langsung Jual (S2)	Entropy	Gain
1.4	AKI-NORMAL and BODY MULUS and MESIN HALUS and CAT KUSAM	3	2	1	0.918296	
	RANGKA					0.251629
	KARAT	1	1	0	0.000000	
	KEROPOS	0	0	0	0.000000	
	MULUS	2	1	1	1.000000	
	AKSESORIS					0.918296
	FULL MODIF	0	0	0	0.000000	
	ORI	1	0	1	0.000000	
	SEMI MODIF	2	2	0	0.000000	



Gambar 2. Pohon Keputusan Hasil Perhitungan Node 1.4

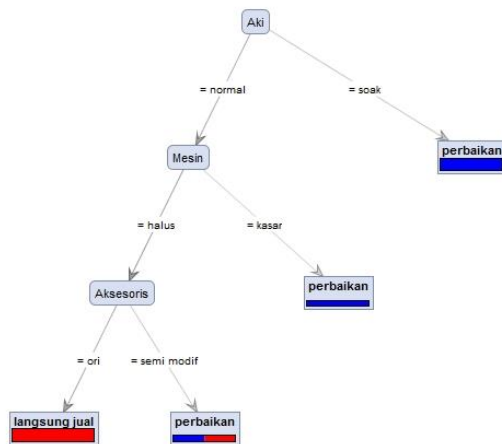
Berdasarkan pohon keputusan terakhir yang terbentuk pada Gambar 2 di atas, maka aturan atau *rule* yang terbentuk adalah sebagai berikut :

1. IF aki = soak, THEN keputusan = perbaikan
2. IF aki = normal AND bodi = lecet, THEN keputusan = langsung jual
3. IF aki = normal AND bodi = penyok, THEN keputusan = perbaikan
4. IF aki = normal AND bodi = mulus AND mesin = kasar, THEN keputusan = perbaikan
5. IF aki = normal AND bodi = mulus AND mesin = halus AND cat = mulus, THEN keputusan = langsung jual
6. IF aki = normal AND bodi = mulus AND mesin = halus AND cat = kusam AND aksesoris = ori, THEN keputusan = langsung jual
7. IF aki = normal AND bodi = mulus AND mesin = halus AND cat = kusam AND aksesoris = semi modif, THEN keputusan = perbaikan

Berdasarkan dari *rule/knowledge* yang dihasilkan terdapat beberapa *rule* yang cukup sesuai dengan kejadian yang terjadi didalam menentukan kelayakan motor bekas, di mana motor bekas yang memiliki kondisi bagus akan langsung jual.

3.3. Implementasi

Implementasi dilakukan dengan menggunakan *software Rapid Miner*. Data yang digunakan dengan format .txt. Hasil dari perhitungan dengan menggunakan *software Rapid Miner* adalah sebagai berikut:



Gambar 3. Tampilan Akhir Decision Tree

Di mana *rule* yang dihasilkan pada gambar 2 adalah :

1. IF Aki = soak THEN keputusan = perbaikan
2. IF Aki = normal AND mesin = kasar THEN keputusan = perbaikan
3. IF Aki = normal AND mesin = halus AND aksesoris = ori THEN keputusan = langsung jual
4. IF Aki = normal AND mesin = halus AND aksesoris = semi modif THEN keputusan = perbaikan

Dari *rule* yang dihasilkan secara umum dapat dilihat bahwa motor bekas yang memiliki kondisi bagus akan bisa langsung dijual, sedangkan motor bekas dengan kondisi yang tidak maksimal bisa dilakukan perbaikan.

Algoritma C4.5 merupakan algoritma yang dianggap membantu dalam melakukan klasifikasi data karena karakteristik data dapat diperoleh dengan jelas, baik dalam bentuk struktur pohon keputusan maupun aturan *if-then*, sehingga memudahkan pengguna dalam melakukan penggalian informasi terhadap data yang bersangkutan.

4. KESIMPULAN DAN SARAN

4.1. Kesimpulan

Berdasarkan uraian penelitian yang telah penulis lakukan maka dapat ditarik kesimpulan :

1. Pemilihan variabel yang akan digunakan dalam memprediksi sangat mempengaruhi *rule* atau *knowledge* yang dihasilkan.

2. Sistem pengklasifikasian tingkat kelayakan motor bekas yang akan dijual menggunakan Algoritma C4.5 dapat digunakan dalam pengambilan keputusan untuk mencari keputusan yang terbaik.

4.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan maka penulis dapat memberikan saran untuk penelitian selanjutnya, sebagai berikut :

1. Perhitungan dengan menggunakan *software Rapid Miner* menghasilkan *rule* yang sedikit berbeda karena ada daun yang hilang meskipun *rule* yang dihasilkan sama. Dari kekurangan tersebut diharapkan nantinya akan dilakukan penelitian lanjutan agar dapat menghasilkan *rule* yang lebih tepat.
2. Pada penelitian ini perlu dilakukan penelitian lanjut untuk membandingkan hasil prediksi dan menentukan teori mana yang menghasilkan prediksi yang baik karena masih ada teknik prediksi lain yang masih bisa digunakan.

DAFTAR PUSTAKA

- A. S. GALATHIYA, A. P. G. A. C. K. B. 2012. Improved Decision Tree Induction Algorithm with Feature Selection, Cross Validation, Model Complexity and Reduced Error Pruning. *International Journal of Computer Science and Information Technologies*, 3(2), 5.
- BRIJESH KUMAR BARADWAJ, S. P. 2011. Mining Educational Data to Analyze Students' Performance. *International Journal of Advanced Computer Science and Applications*, 2, 7.
- HEMLATA SAHU, S. S., SEEMA GONDHALAKAR 2012. A Brief Overview on Data Mining Survey. *International Journal of Computer Technology and Electronics Engineering*, 1, 8.
- IHSAN A. KAREEM, M. G. D. 2014. Improved Accuracy for Decision Tree Algorithm Based on Unsupervised Discretization. *International Journal of Computer Science and Mobile Computing*, 3, 8.
- MARK HALL, E. F., GEOFFREY HOLMES,

- BERNHARD PFAHRINGER PETER REUTEMANN, IAN H. WITTEN 2009. The WEKA Data Mining Software: An Update. 11, 9.
- MUHAMMAD NAEEM AHMED KHAN, S. A. Q. A. N. R. 2013. Gender Classification With Decision Trees. *International Journal of Signal Processing*, 6, 12.
- P.YASODHA, M. K. 2011. Analysis of a Population of Diabetic Patients Databases in Weka Tool. *International Journal of Scientific & Engineering Research*, 2.
- PRIYADHARSINI.C, D. A. S. T. 2014. An Overview of Knowledge Discovery Database and Data mining Techniques. *International Journal of Innovative Research in Computer and Communication Engineering*, 2, 8.
- SEEMA SHARMA, J. A., SANJEEV SHARMA 2013. Classification Through Machine Learning Technique: C4.5 Algorithm based on Various Entropies. *International Journal of Computer Applications*, 82, 8.
- SINGH, S. B. G. N. 2011. THREE PHASE ITERATIVE MODEL OF KDD. *International Journal of Information Technology and Knowledge Management*, 4, 3.
- MINTWAL, S. N. M. K. 2013. Comparison the Various Clustering and Classification Algorithms of WEKA Tools. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3, 13.